图书情报工作
第63卷 第3期 2019年2月



# 基于多特征耦合的数字图书馆知识发现服务优化研究*

张晗[1] 毕强[1] 丁梦晓[1] 李洁[1] 牟冬梅[2]
[1] 吉林大学管理学院 长春 130022 [2] 吉林大学公共卫生学院 长春 130021

**摘要**：[目的/意义]大数据环境下，用户的知识需求由分散向关联转变，利用多特征耦合可以辅助知识发现服务发现资源间的多种相关关系，从而优化知识发现服务。[方法/过程]通过分析文献内部和外部属性特征定义多特征耦合的概念，从功能角度出发，剖析多特征耦合与数字图书馆知识服务之间的关系，结合现有的知识发现系统构建多特征耦合架构，基于数据层－耦合层－服务层三层提出提升知识发现服务供给侧的方法。[结果/结论]数据层保障数据的质量，数据源由单一向混合转变；耦合层提升耦合分析效果，分析单位由粗向细转变，注重细粒度单元间的语义关联；服务层重视用户的交互体验，开发多维可视化功能。

**关键词**：多特征耦合 数字图书馆 知识发现服务
**分类号**：G252 G250.76
**DOI**：10.13266/j.issn.0252－3116.2019.03.002



## 1 引言

信息化、网络化、全媒体的“大数据”时代下的数字图书馆知识发现服务，运用最新的网络工具、理论、技术等，融合关联和协同理念，为用户观测信息、寻找资源、发现知识，并为知识找到用户，从而提供有关知识的智能服务[1]。而大数据驱动下的数据化浪潮促使用户的需求趋向碎片化、个性化、精准化和知识化，用户不再希望浪费时间和精力在海量资源中甄别和判读信息，而是希望直接获取某一具体领域中细粒度的知识、某一主题下精细化的知识体系。并且用户不再以传统的学科方式构建认知，而是基于个人所需的知识点及其知识关联进行智慧活动[2]。因而用户使用数字图书馆检索浏览时，希望检索到的不仅是一篇文献，还有该文献在全部知识体系中的位置及其重要性。假设把一篇文献看作是一个知识点，那么用户除了想找到这个点之外，还想找到该点所在的线以及该线所在的面。同时，大数据也改变了传统的数据分析模式，由假设因果关系到检验样本数据的分析模式转变为实时从大量数据中发现数据间的相关性，以判断事物关联的发展趋势[3]。在现实生活中任何资源都不是孤立的，在其产生过程中都存在着知识的产出与投入的因果关系，这种知识间输出与流入的关系便形成了耦合的基础。耦合是信息资源实体间相互影响而普通存在的一种现象，是指两个及两个以上不同主体建立在同一客体上形成的相互之间的潜在关系，这种潜在的联系是通过两个事物间共同的特征项构成，特征项的量化反映了两者间关联的程度[4]。耦合无形的将资源主体按照其特征单元进行聚类，形成围绕某一主题展开的相关性知识网络，将知识的点、线、面之间进行了有趣的联系，用来揭示点、线、面间隐含的知识和规则，为用户不仅找到了所需的知识，还发现了知识间蕴含的关联和寓意，并且通过潜在的线索挖掘出有价值的情报，辅助科研决策活动。

## 2 多特征耦合和数字图书馆知识发现服务

### 2.1 多特征耦合的定义

耦合的概念最早源自欧洲人 R. M. Fano[5]，直到1963年由麻省理工学院教授 M. Kessler 提出“文献耦合”才在计量学领域广泛受到关注。M. Kessler[6]表示耦合关系是两篇或多篇文献通过共同引用的参考文献

* 本文系国家自然科学基金面上项目“嵌入式知识服务驱动下的领域多维知识库构建”（项目编号：71573102）研究成果之一。
作者简介：张晗（ORCID：0000－0002－4586－0819），硕士研究生，E－mail：15924511205@163.com；毕强（ORCID：0000－0001－7381－4986），教授，博士生导师；丁梦晓（ORCID：0000－0002－6903－8360），硕士研究生；李洁（ORCID：0000－0002－3929－729X），博士研究生；牟冬梅（ORCID：0000－0003－0237－034X，教授，博士生导师，通讯作者，E-mail：moudm@jlu.edu.cn。
收稿日期：2018－07－06 修回日期：2018－10－10 本文起止页码：14－20 本文责任编辑：徐健

张晗，毕强，丁梦晓，等. 基于多特征耦合的数字图书馆知识发现服务优化研究[J]. 图书情报工作,2019,63(3):14－20.

建立起来的,耦合关系的强弱由共同引用的参考文献数量而定,且耦合强度越高,文献间的联系就越紧密,主题相关性也就越高。E. Bassecoulard 等[7]还指出文献耦合关系不仅说明文献主题间有相近的关系,还表明文献作者间的研究主题也有相近的关系。国内最早研究耦合的是王洵[8],他于 1981 年将文献耦合的思想引入中国。随后,邱均平[9]将 M. Kessler 提出的文献耦合概念进一步扩展到作者、期刊、语种、机构、学科之间。张郁蔚[10]还认为文献耦合在提出之时虽然是以文献为耦合单位,但文献是由作者在其研究领域的专业期刊上发表的,而且每种期刊有其主题和所属的类别,所以文献、作者、期刊、主题及所属类别等只是层次上的差别,都可以作为文献耦合的基本单位。随着对耦合关系的研究发展,我们可以利用耦合关系来揭示作者、关键词、正文、期刊、机构等多种特征单元间的相关关系。在信息计量学中耦合分析实质是一种交叉共现的关系[11],通过文献中相同特征的共现关联从而实现不同特征关联的一种分析方法,即从不同维度的特征项出发,建立这些特征项内部特征关联的关系。每一篇文献都是由若干个特征项构成,包括作者、引文、关键词、期刊等,通过对特征项的合理组合构成了文献的主要特征,这也是区别于其他资源的特质[12],而且特征项因其属性不同而各异。因此,这提出的多特征耦合是指融合文献的内部和外部特征,以不同的特征项为主体构建耦合关系,主要以文献本身及其创作者为主;利用不同的特征项为客体,衡量客体间的耦合强度,主要是对关键词、正文内容、参考文献、期刊等进行的定量化分析,从而构建多角度、深层次的耦合关系,为用户揭示丰富全面的知识内容。文献外部特征主要体现在文献的一般性描述元数据中,即不涉及文献的内容,而内部特征是从对文献内容的引用和知识利用角度入手,重点对文献的内容进行分析,如表 1 所示:

**表 1　文献内外部特征分类**

| 特征项 | 属性 | 特征分类 |
|---|---|---|
| 篇名 | 对内容的利用和引用 | 内容特征 |
| 摘要 | | |
| 关键词 | | |
| 正文 | | |
| 作者 | 知识的创造者 | 外部特征 |
| 合著作者 | | |
| 参考文献 | 文献的引用关系 | |
| 文献来源 | 期刊、会议等文献的载体 | |
| 项目资源 | 项目来源 | |
| 收稿日期 | 时间 | |
| 发表日期 | | |

## 2.2　多特征耦合与数字图书馆知识发现服务的关系

在大数据环境下,数字图书馆知识发现服务依托已有的系统平台,通过对结构化、半结构化、非结构化数据的泛在化、语义化、关联化、可视化、智能化的层级解构,挖掘知识单元间潜在的语义关联,并进行知识图谱呈现、个性化定制、智能寻证检索决策等服务应用[13]。彭佳[14]指出目前的知识发现系统正尝试通过关联数据、语义技术等将具有共同特征的资源融合在一起,从语义层面上实现资源的组织聚合,以便用户准确判断多类型资源的内在联系,达到智能化的知识发现。曾建勋[15]对科研文献进行元数据规范化处理及揭示语义关系后,形成了“机构－人员－科研成果”的关系图,对文献中隐含的科研、机构合作等关系充分揭示,实现了基于语义的科研文献推荐和知识发现。可见,知识发现服务正应用关联技术、语义技术等,通过知识的语义关联、引证耦合,朝着智能化、语义化、关联化的方向发展。多特征耦合将外表毫无关联的主体通过客体的相似性特征重新联系在一起,为用户显示出主体间错综复杂的关系,帮助用户厘清各类型主体间的内在关联,挖掘出隐藏在内部的知识,并引导用户发现新的知识结构和规律。多特征耦合为知识发现服务提供从不同视角分析文献资源的关系,以及对这些关系进行深层次的挖掘和重新组织,可为用户提供特定主题检索、个性化知识推荐服务,探寻学科知识结构及演化规律、知识前沿分析与预测等服务。多特征耦合的具体功能见表 2。

一方面,多特征耦合分析从文献内容的不同使用情况出发,通过对文献中共同利用的知识单元进行定量分析,测度知识单元间的耦合强度,根据耦合强度把内容上具有相似性的文献聚在一起形成簇。从而使发现服务可按文献的引用情况将文献组合成小批量的相关集群,从文献利用的角度为用户提供检索途径,并引导用户了解所需的相关文献,扩展用户的阅读范围。基于引用关系建立的耦合,反映的是文献间静态的关系,文献一经发表,其频次便是确定的,不会随时间的变化而变化,通过文献的耦合频次测度当前已发表文献间的相似度,可为用户实现个性化的知识推荐,当用户在查看某篇新刊登的论文时,发现系统会为用户推荐与该文献最相关的、最新的其他论文,还可以帮助用户发现最新的学术信息,尤其是对于需要某领域研究最新信息的用户。同时,发现服务可透过高被引文献的耦合,为用户揭示目前各学科领域活跃的研究前沿,且对不同时期的前沿信息比较,展示各个研究前沿的

表 2 多特征耦合分析的功能

| 主体 | 客体 | 耦合路径 | 耦合功能 |
|---|---|---|---|
| 发表文献 | 参考文献 | 文献—参考文献耦合 | 揭示学科间的发展规律和组织结构,为文献检索提供新途径 |
| 发表文献 | 引文发表时间 | 文献—发表时间 | 探索研究领域内部研究方向,反映研究方向的演化及历史关系 |
| 发文作者 | 文献类别 | 作者—分类号 | 从学科专业角度揭示文献的学科类属,反映作者专业领域的相似性 |
| 发文作者 | 载体 | 作者—期刊 | 揭示领域知识结构,发现领域主题,挖掘共同的学术社区 |
| | | 作者—会议 | |
| 发文作者 | 关键词 | 作者—关键词 | 测度作者相似性,发掘潜在合作关系,识别可能的科研团体,研究主题分布情况 |
| 发文作者 | 正文内容 | 作者—引文 | 揭示学科的智力结构,发现当前研究主题下的活跃群体,挖掘作者在学科发展中的变化轨迹 |
| 发表文献 | 正文内容 | 文献—引文 | 探索知识基础,揭示学科的发展情况,揭示研究前沿热点 |
| 发表文献 | 内容特征词 | 文献—关键词 | 发现文献研究主题,识别核心文献 |
| 引用内容 | 引用内容 | 知识单元—知识单元 | 挖掘领域内在知识结构特征 |



波峰与波谷,方便科研用户在各个研究领域的孕育期和上升初期选择研究切入点,帮助其选择适合的研究方向,使用户可以直观地掌握研究前沿的发展轨迹,了解专业领域研究前沿的焦点,进而探测到潜在的、崭新的研究前沿信息。

另一方面,多特征耦合从文献外部特征角度出发,通过对主体间的相似性要素构建耦合关系。作者是知识资源创新的直接反映,反映的是作者间共同的学术行为,这些行为从侧面反映了作者彼此间学术研究兴趣的联系,基于同一研究主题聚集相关著者。通过同一主题下的著者群,发现服务从某一细分领域向用户揭示学科知识结构,从而引出这一细分领域下学术共同体的智力结构鸟瞰图。而且通过高强度耦合聚集在一起形成的作者群,可从耦合作者群中的著者姓名出发,利用著者检索为用户搜索出某专业下同行所发表的全部有关文献,还可以为用户提供该学科课题研究的定题检索服务。作者间的耦合是一种动态结构,会随着时间的改变而改变。发现服务从作者间动态的耦合关系出发揭示学科领域关注热点的变化以及未来发展趋势,用户在查找资料时,系统可以帮助用户发现最新最常被引用的文献,揭示研究前沿可能发生的变化。而且还可以通过耦合关系将文献按照时间片段聚类,根据耦合强度利用聚类技术把某段时间内相关性高的文献聚集在一起,同一时间段内的文献簇揭示的主题便是该时间内某领域下的知识结构,不同时间段的文献簇则表明知识结构的演化历程。由以上分析可以看出,发现服务可以通过从文献内容利用的不同情况或文献的外部特征出发,构建资源间相关关系,从而为用户提供优质的发现服务。

## 3 数字图书馆知识发现服务多特征耦合架构

目前,数字图书馆知识发现服务都是依靠一定的系统平台展开,如百度学术搜索、超星发现系统、知网KND知识发现网络系统、维普智立方发现系统、万方“淘智”发现系统等主流系统,逐步实现了对学术宏观走向、跨学科交叉及影响和知识再生方向的判断,具备了对特定年代、领域、人物及机构的学术成果态势的多维度对比分析和研究。因此,笔者结合多特征耦合的定义及其与知识发现服务的关系,基于知识发现系统构建知识发现服务多特征耦合架构,该框架自下而上分为数据层、耦合层、服务层 3 个层次,贯穿知识发现系统从初始数据收集处理到满足用户需求的整个过程。见图 1。

(1)数据层:主要负责数据的采集和处理。从不同的数据源中采集数据,获取期刊论文、学位论文、会议论文等资源,在收集数据过程中,数据的字段格式、字段名等不统一现象对耦合分析造成困难。数据层主要分为两个步骤:①对数据预处理,去除噪声数据,获取剩下的蕴含所需知识的数据,这一步是为知识发现系统提供良好知识服务的前提和基础;②对提取之后的数据建立标准的、专业化的知识库,以便后期耦合分析使用。数据层先对无效、缺失、错误、重复等数据进行剔除和清洗,利用数据审计检测数据的质量,转化数据的类型、大小、单位等以方便实现数据的转换和标准化处理,借助归约、排序、抽样等对目标数据实现集成,最终生成相应高质量的引文数据库、关键词、主题词库、作者数据库、作者机构库等专业知识库,便于耦合层准确识别资源的特征项。

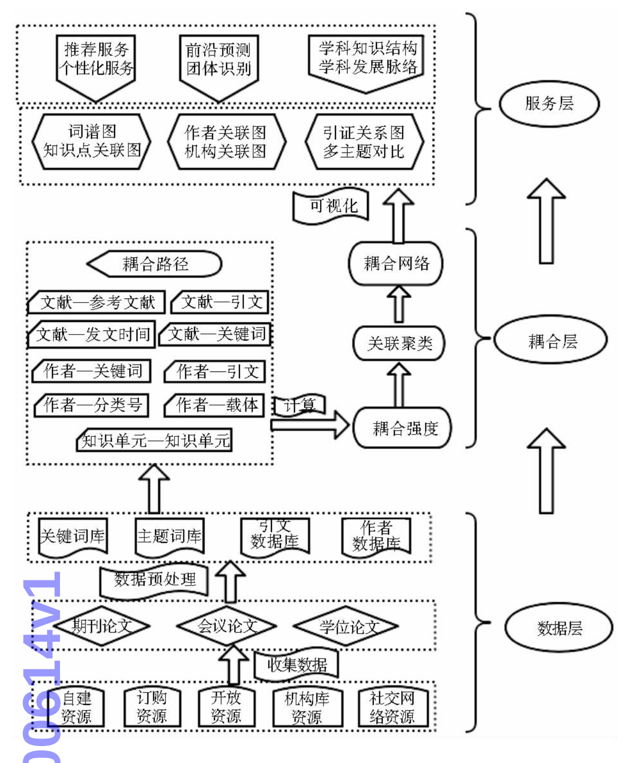


图1 数字图书馆知识发现服务多特征耦合架构

(2)耦合层：主要负责多特征耦合关系的生成和关联聚类。多特征耦合关系是通过两个主体间的相似性要素构建，并通过客体间的相关性程度获得，相关性大小用耦合强度来衡量。耦合强度越大，相似性程度越高，耦合网络中的聚类效果越佳。首先，利用已构建好的知识库，对初步优化好的目标数据通过文献计量分析中的共词分析、共篇分析、引文分析等方法对资源的不同特征项进行分析，识别主体间拥有的相似要素，构建不同主体资源与客体特征的耦合路径，如从文献内容的不同使用情况出发构建文献-引文、文献-关键词或作者-引文、作者-关键词间的耦合路径，通过衡量客体间耦合强度从而获得主体间相似性的大小，集成聚类强的关联节点，构建具备紧密联系的知识链，利用中心性分析、凝聚子群分析、核心-边缘分析、时序分析等方法构建耦合知识网络[16]，通过网络中的节点距离和密度反应节点间联系的亲疏程度，从而识别出领域的核心作者及其学术影响力、潜在学术团体的凝聚规模、学科主题关系的交叉程度、学科发展的历史和趋势及知识间的交流和流动程度等。这一层中，主要是对选取的耦合单位建立耦合关系，形成紧密关联的耦合网络，借助可视化技术对耦合网络进行静动态的图形表征，为用户呈现目标检索、解读、预测等具备视觉表达的知识域。

(3)服务层：主要负责连接用户与知识发现系统，是用户查询输入与系统结果输出的纽带。目前，知识发现系统可根据用户的检索策略、行为等数据，在耦合分析层基础上通过可视化图形形象地展示分析结果，为用户提供词谱图、知识点关联图、作者关联图、机构关联图、多主体对比及引证关系图等，实现对主题词、关键词、知识点、作者、作者机构等的立体关联，挖掘和追溯领域研究主题，分析多领域、多主题、多学科等的学术研究的生长方向，预测学术发展现状及未来趋势等，为用户解读知识脉络、分析多维知识、智能化决策等提供了辅助途径。同时通过高耦合强度资源的聚类，根据聚类结果向用户提供资源推荐和知识服务，引导用户需求，协助展开个性化服务。在这一层中，数字图书馆知识发现系统数据的清理更新，相关数据库的完善、耦合关系的全面分析、知识关联图谱的动态升级等都会影响服务质量和效果。

## 4 数字图书馆知识发现服务多特征耦合优化

从前文提出的数字图书馆知识发现服务多特征耦合框架可知，目前的知识发现系统拥有海量的各类型资源及专业规范的知识库，通过对馆藏资源进行纵横结合、深达知识内在关系的挖掘和分析，揭示知识间的复杂交叉关系，大尺度、多维度的对特定领域、年代、著者及机构的学术研究成果进行对比分析和研究，并以可视化方式展示。虽然这些知识发现系统已能辅助用户实现智能检索、个性化推荐、学术科研辅助等服务，但随着信息环境的变化，用户的需求也发生了变化，对知识发现服务也有了更高的期待。在大数据环境下，用户对知识获取的需求由粗转细、由分散向综合转变，针对性、精细化的需求特征明显，而且用户对发现服务的要求更倾向于资源的深度和广度分析、知识服务的竞争力分析、知识创新性的预测、综述服务的发展态势等高附加值的服务[17]。同时，用户希望数字图书馆提供的是一个交互友好、关联协作的发现平台，引导并支持其发现未知的知识结构和规律。因此，面向用户不断变化的需求，数字图书馆知识发现服务应紧紧把握住“环境改变需求，需求变革服务”的理念，提升服务供给侧的能力，通过数据层-耦合层-服务层的优化实现资源间的更有效关联，促进重要的资源发现和重用，帮助用户创造新知识、解决新问题。

### 4.1 数据层注重由单一数据源向混合数据源转变

合理选择耦合单位的来源是多特征耦合分析的出发点，也是保证耦合分析的质量基础，才能为用户发现真正所需的信息，提供针对性的发现服务。进行耦合分析的第一步是对数据层中资源质量的把关，尤其各类型资源中作者姓名歧义性的问题，一旦处理不正确，作者和其对应的信息不能完全匹配，缺乏足够的信息来定位和识别作者，导致用户检索数据时往往会检索到不相干的同名作者发表在其他领域的文献，一旦这些数据作为耦合分析的数据对象，那么获取的数据量将会增大，使得后期耦合分析工作产生额外的负担。因此，数据层要注重有效的数据消歧处理，为耦合分析获取可靠的数据做准备，实现其在整个分析流程中的价值。知识发现系统从数据源头对作者姓名进行规范控制，主要采取对作者全名备注、对作者及作者机构信息、联系信息等联合标引等方法对不同作者身份进行甄别[18]，帮助用户准确定位数据。还有一些从作者源头确认身份的辅助措施，例如利用 ResearchID、ThuRID、指纹识别等新兴技术可以在客观上为作者消歧提供帮助。

在大数据环境下，数字图书馆不仅包括传统的数字化资源，如图书、期刊、报纸、学位论文等，还包括半新资源，如免费网上资源、文献型电子资源、科学数据、关联性标注资源等，还有大量的全新资源，如用户的行为数据、评论数据、日志数据、地理位置数据等。因此，知识发现服务的耦合分析由单一的数据源向混合数据源转变，不应只局限于已发表的期刊论文、会议论文，还应综合利用蕴含近期和远期研究思想的科技规划文本、科研项目申请书、学术记录、热点话题报告、用户发表的评论等数据，作为构建耦合分析数据库的基础。单一数据源的耦合分析是片面的，难以满足用户全方位的信息需求，尤其是对一些问题模糊、知识不清的弱信息的需求。利用混合数据源进行耦合分析，可以通过多种类型的资源建立耦合关系，挖掘多种资源间的联系，形成多层次的耦合网络，揭示不同类型资源间同质层节点的关联，反映异质层节点间的联系，每一层所展示的内容都不尽相同[19]，而且网络层次越多，关联越深入，愈能清晰地梳理问题领域的知识点和结构，准确鉴别问题和问题间的关系，为用户揭示丰富的隐性知识，满足用户对信息的多方位需求。

### 4.2 耦合层注重由粗粒度文本分析向细粒度语义关联分析转变

在大数据环境下，由于发布数据的成本变得低廉，发布渠道更加多样，用户无论是在数字图书馆内部还是外部都面临着海量的碎片化资源，用户很难完整获取和阅读自己所需领域的全部相关资源，驱使用户对资源的获取更加倾向于精细化、语义化和关联化。而且目前的知识发现系统都是基于文献中的相同点建立耦合关系，很少从语义层面实现关联，而基于语义层面的关联能克服资源间的语义异构，实现资源的广度关联和深度聚合。因此，知识发现系统在选择耦合单位时要从粗粒度的知识单元深入到细粒度的知识单元中，耦合的控制单位由粗转细，文献中包含的大量细粒度知识单元及其潜在的语义结构将会产生知识增值，促进对新知识的创造。挖掘到的知识单元颗粒度越小，代表的知识越精确，通过具有语义关联的细粒度知识单元进行耦合分析，可以挖掘出知识单元间的潜在关联。文献中知识单元出现的位置不同对主题的表达能力也不同，对文献中涉及表达主题的标题、摘要、关键词、增补关键词等凝练部分，全文的前沿、讨论和结论部分，以及文章段落的位置、句子在段落中的位置和词在句子中的位置，综合考虑语境进行细粒度的挖掘和划分[20]，还可以借鉴 CNKI 知识元搜索，该功能通过对文献的统计信息、表格、图片等信息的获取基础上，进一步挖掘到文献的数据事实、创新点等精细粒度的知识单元[21]。利用语义分析技术实现资源的精细粒度挖掘和基于语义的纵横关联，为用户揭示出研究领域的结构细节，对研究领域内容进行详细的解读，使得数字图书馆知识发现服务从传统的用户需求驱动转变为战略规划和决策服务，提升数字图书馆决策辅助的精准化程度。

### 4.3 服务层注重多维可视化的交互开发

发现系统的服务是在多源数据融合、精细粒度知识单元语义化分析和广度关联基础上展开，其结果的展示影响着其服务效果的发挥，同时也对可视化性能提出了更高的要求。大数据环境下用户希望通过良性的可视化交互感知，对海量数据进行分析，从而获取知识，内化为智慧，做出决策[22]。因此，数字图书馆知识发现服务应借鉴数据可视化技术（综合计算机图形学、图像处理、人机交互等技术），对海量、异构、多来源的数据进行整合分析，利用图形元素对多源数据库中的

每个数据项表示,通过图形、符号、颜色、纹理等表达方式及时生成可视化方案,而且以多维数据的形式表示数据的各个属性值,并对其进行切片、块、旋转等动作的剖析,从多维度、全方位观察数据,从数据中抽取出信息,挖掘出信息间的潜在规律,分析出有意义的知识模式,帮助用户发现新的知识,并以故事的形式将知识传递给用户,使用户在获取信息之外还有所启发。同时,为了向用户进一步反映出文献、作者、学科等隶属国家和地区间的分布情况、交流合作程度以及知识流动状况,也为了向更多科研用户形象地展示出不同国家和地区科研水平的高低和研究侧重点,发现系统可借助地理空间分析方法,通过嵌入 Google maps 或 Yahoo! 地理编码[23],创造带有地理位置信息的可视化图谱,提升可视化分析与预测的科学性。通过提升可视化的性能,为用户提供良好的交互情境,引导用户获得有价值的知识,辅助用户进行学科溯源、前沿掌握、核心知识识别等。

## 5 结语

在大数据环境下,用户已不再为了获取更多的资源而使用数字图书馆,不是由因到果寻找知识,而是更加注重知识间的相关关系、相互关系,使得数字图书馆对资源内容及其关系的发现愈发重要,提供知识发现、知识关联图谱、个性化推荐、特性检索、辅助科研决策等服务愈发急迫。笔者通过定义多特征耦合构建了知识点、线、面间的联系,基于多特征耦合在数字图书馆知识发现服务中发挥的功能,分析二者之间的关系,在现有知识发现系统的基础上建立了多特征耦合的架构,面向大数据环境下用户知识需求的演化,进而从架构的数据层、耦合层和服务层提出相应的优化建议,以提升数字图书馆知识发现服务的效果。



**参考文献:**

[1] 卢章平,袁润,王正兴. 发现服务:大学与研究机构图书馆的趋势[J]. 中国图书馆学报,2014,40(3):20－26.
[2] 夏立新,白阳,张心怡. 融合与重构:智慧图书馆发展新形态[J]. 中国图书馆学报,2018,44(1):35－49.
[3] 舍恩伯格,库克耶. 大数据时代:生活、工作与思维的大变革[M]. 盛杨燕,周涛,译. 杭州:浙江人民出版社,2013.
[4] 邱均平. 信息计量学[M]. 武汉:武汉大学出版社,2007.
[5] EGGHE L, ROUSSEAU R. Cocitation, bibliographic coupling and a characterization of lattice citation networks[J]. Scientometrics, 2002, 55(3):349－361.
[6] KESSLER M. Bibliographic coupling between scientific papers[J]. American documentation,1963(14):123－131.
[7] BAEERCOULARD E, LELU A, ZITT M. Mapping none sciences by citation flows: a preliminary analysis[J]. Scaentometrics, 2007,70(3):859－880.
[8] 王洵. 开斯勒与"文献合配"[J]. 情报科学,1981,2(4):29－32.
[9] 邱均平. 论"引文耦合"与"同被引"[J]. 图书馆,1987(3):13－19.
[10] 张郁蔚. 以直接引用、书目耦合及共同作者探讨图书资讯学跨学科之变迁[D]. 台北:台湾大学,2009.
[11] 邱均平,王菲菲. 基于共现与耦合的馆藏文献资源深度聚合研究探析[J]. 中国图书馆学报,2013,39(3):25－33.
[12] MORRIS S A. Unified mathematical treatment of complex cascaded bipartite networks: the case of collections of journal papers[D]. Stillwater:Oklahoma State University. 2005.
[13] 李洁,毕强,张晗,等. 数据驱动下数字图书馆知识发现的服务研究[J]. 情报资料工作,2018(4):6－14.
[14] 彭佳,郑巧英. 信息资源聚合与组织研究——以发现系统为例[J]. 图书馆杂志,2016,35(3):80－85.
[15] 曾建勋,丁遒劲. 基于语义的国家科技信息发现服务体系研究[J]. 中国图书馆学报,2017,43(4):51－62.
[16] 崔金栋,孙遥遥,郑鹊,等. 我国政务微博社会网络特征对比分析实证研究[J]. 情报科学,2016,34(12):120－126,132.
[17] 毕强,闫晶,李洁. 大数据时代数字图书馆服务转型面临的新形势与新要求[J]. 情报理论与实践,2017,40(12):12－16,5.
[18] 董克. 数字文献资源多元深度聚合研究[D]. 武汉:武汉大学,2014.
[19] 郭秋萍,梁梦丽,刘秀丽,等. 基于作者－关键词－引文多重共现的超网络知识关联研究[J]. 情报理论与实践,2016,39(7):20－26.
[20] 任红娟. 一种内容和引用特征融合的知识结构划分方法研究[J]. 中国图书馆学报,2013,39(5):76－82.
[21] 王福. 基于信息资源组织视觉的新型 OPAC 系统设计研究[J]. 图书馆工作与研究,2015(3):38－40.
[22] 陈为,沈则潜,陶煜波. 数据可视化[M]. 北京:电子工业出版社,2013.
[23] RICHARD K, KEVIN W B. Using global mapping to create more accurate document-level maps of research fields[J]. Journal of the American Society for Information Science and Technology,2011,62(1):1－18.

**作者贡献说明:**

张晗:提出研究思路,撰写论文;
毕强:提出论文框架,完善论文思路;
丁梦晓:英文摘要撰写,修改论文;
李洁:完善研究思路,修改论文;
牟冬梅:完善研究思路,修改论文。

## Research on Knowledge Discovery Service Optimization of Digital Library Based on Multi-feature Coupling

Zhang Han[1] Bi Qiang[1] Ding Mengxiao[1] Li Jie[1] Mu Dongmei[2]

[1] School of Management, Jilin University, Changchun 130022

[2] School of Public Health, Jilin University, Changchun 130021

**Abstract:** [**Purpose/significance**] In the context of big data, the user's knowledge needs are changed from decentralization to correlation, and multi-feature coupling is used to assist the knowledge discovery service to discover multiple correlations between resources, thereby optimizing knowledge discovery services. [**Method/process**] The concept of multi-feature coupling was defined by analyzing the internal and external attribute characteristics of the literature. This paper analyzed the relationship between multi-feature coupling and digital library knowledge service according to the function of multi-feature coupling. Then, by combining the existing knowledge discovery system, the multi-feature coupling structure was constructed. And the method of improving the supply side of the knowledge discovery service was proposed based on data layer-coupling layer-service layer. [**Result/conclusion**] The data layer guarantees the quality of the data, the data source changes from single to mixed; the coupling layer enhances the effect of coupling analysis, the unit of analysis changes from coarse-grained to fine-grained, the semantic association between fine-grained units attracts much attention; the service layer attaches importance to the user's interactive experience and develops multi-dimensional visualization function.

**Keywords:** multi-feature coupling digital library knowledge discovery service